Journal of Multivariate Analysis **82**, 422–430 (2002)
doi:10.1006/jmva.2001.2031

A Formula for the Tail Probability of a Multivariate Normal Distribution and Its Applications[1]

Jürg Hüsler

Universität Bern, Bern, Switzerland
E-mail: juerg.huesler@stat.unibe.ch

and

Regina Y. Liu and Kesar Singh

Rutgers University
E-mail: rliu@stat.rutgers.edu, kesar@stat.rutgers.edu

An exact asymptotic formula for the tail probability of a multivariate normal distribution is derived. This formula is applied to establish two asymptotic results for the maximum deviation from the mean: the weak convergence to the Gumbel distribution of a normalized maximum deviation and the precise almost sure rate of growth of the maximum deviation. The latter result gives rise to a diagnostic tool for checking multivariate normality by a simple graph in the plane. Some simulation results are presented. © 2002 Elsevier Science (USA)

AMS 1991 subject classifications: 60G70; 62H05; 62J20.

Key words and phrases: multivariate normal distribution; tail probability; Gumbel distribution; maximum deviation; growth rate; sum of χ^2 random variables.

1. INTRODUCTION AND MAIN RESULTS

Let $X, X_1, \ldots$ be a sequence of i.i.d. random variables in $\mathbb{R}^d$, each following a normal distribution with mean μ and covariance matrix Σ, possibly singular. We denote this distribution by $N(\mu, \Sigma)$ and by

$$M_n = \max_{1 \leqslant i \leqslant n} \|X_i - \mu\|$$

[1] Research supported in part by the Federal Aviation Administration and the National Science Foundation. The authors thank Dr. Jesse Parelius for his assistance in preparing the graphs.

0047-259X/02 $35.00

the maximum deviation of $\{X_1, \ldots, X_n\}$ from the mean. Here $\|x\|$ stands for the Euclidean norm of $x \in \mathbb{R}^d$.

This paper has two goals. The first is to establish an exact asymptotic formula for the tail probability of the random variable $\|X-\mu\|$. It is a formula for the tail probability of a finite sum of weighted, independent χ^2-distributed random variables. The second is to derive from this formula two asymptotic results for the maximum deviation from the mean, namely the weak convergence to the Gumbel distribution of a normalized maximum deviation and the exact almost sure rate of growth of the maximum deviation. The rate of growth gives rise to a graphical tool for checking multivariate normality. This graphical diagnostic tool is illustrated in some simulation examples.

Let $\{\lambda_i, i=1, \ldots, d\}$ be the eigenvalues of Σ, written in decreasing order where

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_d.$$

Let $\lambda_{\max}$ be the maximum eigenvalue, i.e., $\lambda_{\max} \equiv \lambda_1$, and m the multiplicity of $\lambda_{\max}$, i.e., $m = \#\{j: \lambda_j = \lambda_{\max}, 1 \leqslant j \leqslant d\}$. Denote by $\Gamma(s) = \int_0^\infty t^{s-1} e^{-t}\, dt$ the gamma function.

THEOREM 1. *Assume that X follows $N(\mu, \Sigma)$, where Σ is allowed to be singular. Then*

$$P(\|X-\mu\| > \sqrt{y\lambda_{\max}}\,) \sim \left(2^{1-\frac{m}{2}} \Big/ \Gamma\left(\frac{m}{2}\right)\right) e^{-\frac{y}{2}} y^{\frac{m}{2}-1} \prod_{j=m+1}^{d} (1-\lambda_j/\lambda_{\max})^{-1/2} \tag{1}$$

as $y \to \infty$.

Using this result, we derive the limit distribution for M_n. The derivation is based on the following linear normalization scheme which is standard in the analysis of extreme values (see, for example, Leadbetter *et al.* (1983), Resnick (1987), Falk *et al.* (1994), or Embrechts *et al.* (1997)).

Let $u_n(x) = b_n + a_n x$, $x \in \mathbb{R}$, where

$$\begin{aligned} a_n &= \sqrt{\lambda_{\max}/(2\log n)}\,, \qquad \text{and} \\ b_n &= [\lambda_{\max}(2\log n + 2\log c_m - 2\log\Gamma(m/2) + (m-2)\log\log n)]^{1/2}, \end{aligned} \tag{2}$$

with $c_m = \prod_{j=m+1}^d (1-\lambda_j/\lambda_{\max})^{-1/2}$. Theorem 2 below shows that the normalized maximum $(M_n - b_n)/a_n$ converges weakly to the Gumbel distribution.

THEOREM 2. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables following the d-dimensional normal distribution $N(\mu, \Sigma)$, where the covariance matrix Σ is allowed to be singular. Then*

$$P\{M_n \leqslant b_n + a_n x\} \xrightarrow{d} G(x) = \exp(-e^{-x}) \tag{3}$$

as $n \to \infty$. Here a_n and b_n are specified in (2), and "$\xrightarrow{d}$" denotes convergence in distribution.

Combined with Lemma 1 below, Theorem 1 also gives the following almost sure growth rate of M_n as $n \to \infty$.

THEOREM 3. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, each following the d-dimensional normal distribution $N(\mu, \Sigma)$, where Σ is allowed to be singular. Then M_n grows almost surely at the exact rate of $\{2\lambda_{\max}(\log n + \frac{m}{2} \log\log n)\}^{1/2}$, meaning*

$$P(\max_{1 \leqslant i \leqslant n} \|X_i - \mu\| \geqslant \{2\lambda_{\max}(\log n + s \log\log n)\}^{1/2}, \text{i.o.}) = \begin{cases} 1 & \text{if} \quad s \leqslant \dfrac{m}{2}, \\ 0 & \text{if} \quad s > \dfrac{m}{2}. \end{cases} \tag{4}$$

Remark 3.1. The phrase "grows at the exact rate" in Theorem 3 means that the growth rate is exact since $m/2$ is the exact cut-off point for the constant s in the left-hand side of the expression and there are no other unspecified constants which need to be estimated.

Clearly, if $\Sigma = \mathbf{I}$, then $\lambda_{\max} = 1$ and $m = d$. In this case, it is interesting to note that the dimension of the data appears only in the secondary term involving $(\log\log n)$ in the rate of growth.

Remark 3.2. The result in Theorem 3 can be used as a diagnostic tool for validating the assumption of multivariate normality. In other words, by comparing the growth curve given in (4) with the graph of $\{\max_{1 \leqslant i \leqslant n} \|X_i - \mu\|, n = 1, \ldots\}$ from the observed sample, we can visually determine whether or not the sample $\{X_1, \ldots, X_n\}$ is drawn from a multivariate normal distribution. If the two curves are far apart from each other, then it indicates that the observed sample does not follow a normal distribution. Some simulated results are presented in Figs. 1 to 3. Figure 1 contains the plot (expressed in a solid curve) of the maximum of a random sample from the bivariate standard normal distribution as n grows to 5000, against the growth curve (expressed in a dotted curve) given in Theorem 3. The same set of curves for a sample from the standard 10-dimensional normal

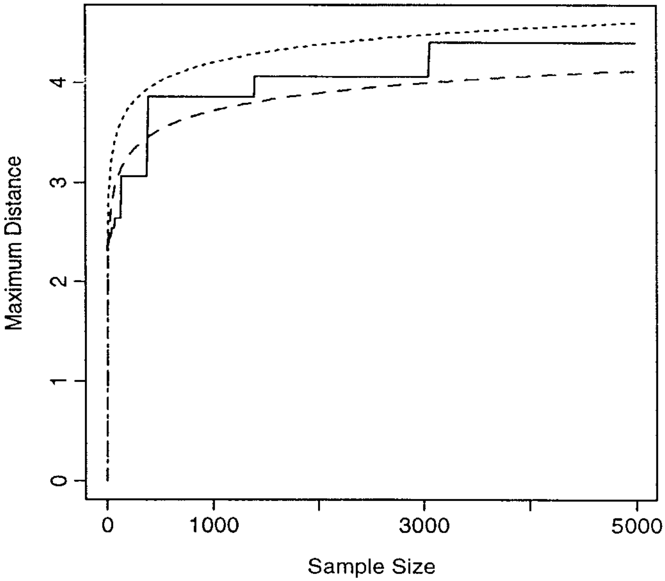

FIG. 1. Maxima of normal samples ($\mathbb{R}^2$).

distribution are given in Fig. 2. In both cases, the two curves appear to match closely. The lower dashed curve in each figure is the growth curve in Theorem 3 without the secondary ($\log \log n$) term. Comparing with the solid curves, the lack of fit of the lower dashed curve becomes more apparent as the dimension of the underlying distribution increases. This shows the crucial role of the secondary term in Theorem 3. Finally, the same three curves are plotted in Fig. 3 for a bivariate exponential sample, where the lack of fit of the growth curve intended for the normal case is evident.

Remark 3.3. Before applying the growth rate curve to verifying the normality of a distribution, we recommend that the observed data be standardized (or sphericized) first, assuming that the covariance matrix Σ is nonsingular. This forces the identity matrix $\mathbf{I}$ to be the covariance matrix of the sphericized data and thus avoids the problem of determining the multiplicity of $\lambda_{\max}$ from an estimator $\hat{\Sigma}$. In the process of sphericizing data, we should use a robust version of $\hat{\Sigma}$. A simple approach has been suggested in Liu *et al.* (1999), which uses a normalized covariance matrix estimated from only the 50% most central data points. The *centrality* of a data point with respect to the given sample is measured by a notion of data depth, as described in Section 2 of the same paper. Theorem 8.1 in Liu *et al.* (1999) shows that $\Sigma(0.5) = \eta(0.5) \times \Sigma$, where $\Sigma(0.5)$ denotes the covariance matrix

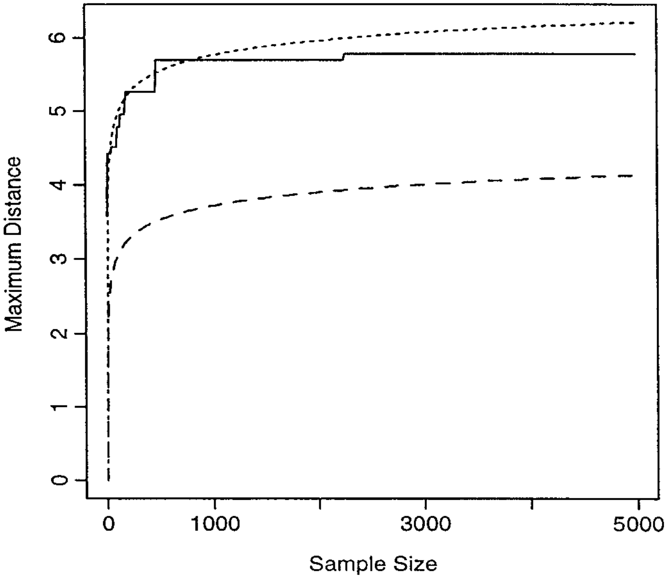

FIG. 2. Maxima of normal samples ($\mathbb{R}^{10}$).

when the population distribution is conditional on the central 50% region, and

$$\eta(p) = \frac{E(R^2 \mid R^2 \leqslant \xi_p)}{E(R^2)}.$$

Here $R^2 = (X-\mu)\,\Sigma^{-1}(X-\mu)'$, and ξ_p stands for the pth quantile of R^2 which is χ^2-distributed with d degrees of freedom. The density function of R^2 is given by

$$g(r^2) = \frac{(2\pi)^{-d/2}\,\pi^{d/2}}{\Gamma\left(\dfrac{d}{2}\right)} \cdot (r^2)^{d/2-1}\, e^{-r^2/2}.$$

For example, in the case of a bivariate normal distribution, $d = 2$ and the distribution of R^2 is exponential with mean 2. Thus

$$\begin{aligned}\eta(0.5) &= 2\int_0^{\ln 2} ye^{-y}\,dy \Big/ \int_0^{\infty} ye^{-y}\,dy \\ &= (1-\ln 2) \approx 0.31.\end{aligned}$$

This suggests that $\{\hat{\Sigma}(0.5)/0.31\}$ should be a reasonable choice for the purpose of standardizing the given bivariate sample.

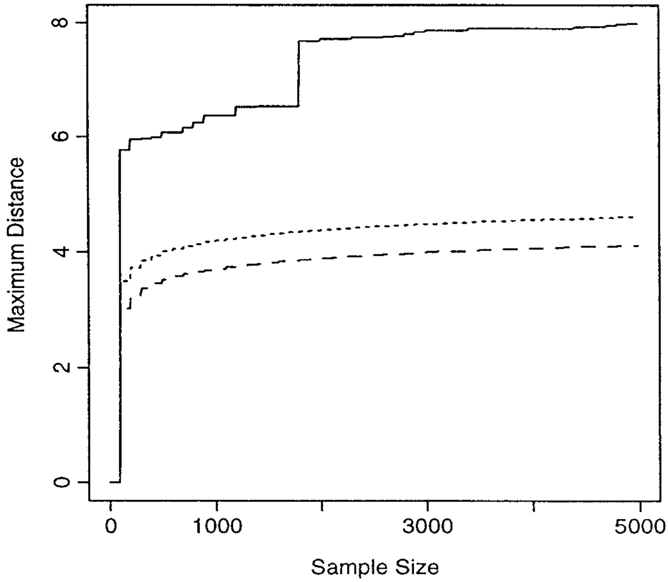

FIG. 3. Maxima of exponential samples ($\mathbb{R}^2$).

2. PROOFS

Proof of Theorem 1. Without loss of generality, we assume $\mu = 0$. We can express X in terms of its principal components $\{\sqrt{\lambda_j}\, U_j, j = 1, ..., d\}$, where U_j are i.i.d. standard normal random variables,

$$\|X\|^2 = \sum_{j=1}^{d} \lambda_j U_j^2.$$

This can be rewritten as

$$\|X\|^2 = \lambda_{\max} \left\{ W + \sum_{j=m+1}^{d} \tau_j U_j^2 \right\},$$

where $W \equiv \sum_{j=1}^{m} \lambda_j U_j^2 / \lambda_{\max} = \sum_{j=1}^{m} U_j^2$, and $\tau_j = \lambda_j / \lambda_{\max}$. Note that W follows a χ^2-distribution with m degrees of freedom and that $\tau_j < 1$, for all $j > m$. We use an iterative method to derive the tail probability of the distribution of $\{W + \sum_{j=m+1}^{d} \tau_j U_j^2\}$.

(I) Consider first the random variable $(W + \tau V)$, where τ is a constant with $1 > \tau > 0$ and V follows a χ^2-distribution with k degrees of freedom, with density function $f_V(x)$. We begin by examining the tail behavior of $P\{W+\tau V>y\}$ where $\tau<1-\varepsilon<1$ for some $\varepsilon>0$. Note that

$$\begin{aligned} P\{W+\tau V>y\} &= \int_0^\infty P\{W>y-\tau x\}\, f_V(x)\,dx \\ &= \int_0^{y(1-\varepsilon)/\tau} P\{W>y-\tau x\}\, f_V(x)\,dx \\ &\quad + \int_{y(1-\varepsilon)/\tau}^\infty P\{W>y-\tau x\}\, f_V(x)\,dx. \end{aligned} \tag{5}$$

To evaluate the second integral above, we use the following approximation as $y\to\infty$,

$$P\{V>y\} \sim \exp(-y/2)\, y^{k/2-1}\gamma_k, \tag{6}$$

where $\gamma_k=(2^{k/2-1}\Gamma(k/2))^{-1}$ is the normalization of the χ^2-distribution. In view of (6), the second integral in (5) is bounded by $P\{V>y(1-\varepsilon)/\tau\}$, and thus is $o(P\{W>y\})$ since $(1-\varepsilon)/\tau>1$. As for the first integral in (5), we divide its range into two parts, $[0, x_0]$ and $[x_0, y(1-\varepsilon)/\tau]$, for some large x_0 and y sufficiently large. Then the integral over $[0, x_0]$ can be approximated by

$$\begin{aligned} &\int_0^{x_0} P\{W>y-\tau x\}\, f_V(x)\,dx \\ &\sim \int_0^{x_0} e^{-y/2+\tau x/2}\,(y-\tau x)^{m/2-1}\,\gamma_m\, e^{-x/2}\, x^{k/2-1}\,\gamma_k\,dx \\ &= e^{-y/2}\, y^{m/2-1}\,\gamma_m \int_0^{x_0} (1-\tau x/y)^{m/2-1}\, e^{-x(1-\tau)/2}\, x^{k/2-1}\,\gamma_k\,dx \\ &\sim P\{W>y\}(1-\tau)^{-k/2}\, P\{V\leqslant x_0(1-\tau)\} \end{aligned}$$

as $y\to\infty$. Note that the last factor $P\{V\leqslant x_0(1-\tau)\}\to 1$ as $x_0\to\infty$. Similarly, we can approximate the integral over $[x_0, y(1-\varepsilon)/\tau]$ by using (6) for $P\{W>y-\tau x\}$ since $y-\tau x\geqslant \varepsilon y\to\infty$

$$\begin{aligned} &\int_{x_0}^{y(1-\varepsilon)/\tau} P\{W>y-\tau x\}\, f_V(x)\,dx \\ &\quad = O(P\{W>y\}) \int_{x_0}^{y(1-\varepsilon)/\tau} e^{-x(1-\tau)/2}\, x^{k/2-1}\,\gamma_k\,dx \\ &\quad = o(P\{W>y\}) \end{aligned}$$

as $x_0 \to \infty$. Hence

$$P\{W+\tau V > y\} \sim (1-\tau)^{-k/2}\, P\{W > y\}$$
$$\sim (1-\tau)^{-k/2}\, e^{-y/2}\, y^{m/2-1}\, \gamma_m. \tag{7}$$

If $\tau = 0$, the result follows immediately from (6).

(II) The above approximation can be generalized to the case $P\{W + \tau_1 V_1 + \tau_2 V_2 > y\}$, where W, V_1, and V_2 are independent χ^2-distributed random variables with respectively m, k_1 and k_2 degrees of freedom. The asymptotically relevant part is the integral on $[0, x_0]$ as above in part (I). Using the result (7), this can be approximated in the same way as in (I) and leads to

$$\int_0^{x_0} P\{W+\tau_1 V_1 > y-\tau_2 x\}\, f_{V_2}(x)\, dx$$
$$\sim (1-\tau_1)^{-k_1/2}\, e^{-y/2}\, y^{m/2-1}\, \gamma_m\, (1-\tau_2)^{-k_2/2}\, P\{V_2 \leqslant x_0(1-\tau_2)\}$$
$$\sim (1-\tau_1)^{-k_1/2}\, e^{-y/2} y^{m/2-1} \gamma_m\, (1-\tau_2)^{-k_2/2}$$

as $y \to \infty$ and $x_0 \to \infty$. Iterating the procedure, we obtain

$$P\left\{W + \sum_{j=m+1}^{d} \tau_j U_j^2 > y\right\} \sim \prod_{j=m+1}^{d} (1-\tau_j)^{-1/2}\, e^{-y/2}\, y^{m/2-1}\, \gamma_m$$

which is our statement (1). The above argument holds even if some of the τ_j's are equal to 0. ∎

Proof of Theorem 2. The desired statement (3) is equivalent to

$$nP\{\|X-\mu\| > b_n + a_n x\} \to e^{-x} \tag{8}$$

for any $x \in \mathbb{R}$. Since $b_n \sim \sqrt{2\lambda_{\max} \log n}$, we have

$$u_n^2(x) = b_n^2 + 2\lambda_{\max} x(1+o(1)).$$

Thus,

$$u_n^2(x)/\lambda_{\max} = 2 \log n + 2 \log c_m - 2 \log \Gamma(m/2)$$
$$+ (m-2) \log \log n + 2x + o(1).$$

Using Theorem 1, we obtain

$$nP\{\|X-\mu\| > b_n + a_n x\}$$
$$\sim n \exp\{-u_n^2(x)/(2\lambda_{\max})\}(u_n^2(x)/\lambda_{\max})^{m/2-1}\, c_m/(2^{m/2-1}\Gamma(m/2))$$

$$\sim n\exp\{-\log n - \log c_m + \log \Gamma(m/2) - (m/2-1)\log\log n - x + o(1)\}$$
$$\times (2\log n)^{m/2-1}\, c_m/(2^{m/2-1}\Gamma(m/2))$$
$$\sim \exp(-x), \qquad \text{as} \quad n\to\infty,$$

and (8) is established. ∎

Proof of Theorem 3. The proof requires the following Lemma 1, which allows us to examine the partial maxima of a sequence of random variables by monitoring the i.i.d. sequence itself. This in turn allows us to apply the Borel–Cantelli lemma to establish an almost sure statement. The result in Theorem 3 then follows from Theorem 1. ∎

LEMMA 1. *Let $\{v_n\}$ and $\{c_n\}$ be two real-valued sequences. Assume that $\{c_n\}$ increases to ∞. Then,*

$$\{v_n \geqslant c_n, \text{ i.o.}\} \qquad \textit{if and only if} \quad \{\max_{1\leqslant i\leqslant n} v_i \geqslant c_n, \text{ i.o.}\}.$$

Proof. We only need to prove the direction "if", since the other direction is straightforward. Assume the contrary: $v_n \geqslant c_n$, finitely often. Then there is some m such that $v_n < c_n$, for all $n \geqslant m$. Hence, $\max_{m\leqslant i\leqslant n} v_i \leqslant c_n$, for $n \geqslant m$. Since for n sufficiently large $\max_{1\leqslant i\leqslant m} v_i < c_n$, it follows that $\max_{1\leqslant i\leqslant n} v_i \geqslant c_n$ finitely often. ∎

REFERENCES

P. Embrechts, C. Klüppelberg, and Th. Mikosch, "Modelling Extremal Events," Springer-Verlag, Berlin, 1997.

M. Falk, J. Hüsler, and R. D. Reiss, "Laws of Small Numbers: Extremes and Rare Events," Birkhäuser, Basel, 1994.

M. R. Leadbetter, G. Lindgren, and H. Rootzén, "Extremes and Related Properties of Random Sequences and Processes," Springer-Verlag, Berlin, 1983.

R. Liu, J. Parelius, and K. Singh, Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion), *Ann. Statist.* **27** (1999), 783–858.

S. I. Resnick, "Extreme Values, Regular Variation, and Point Processes," Springer-Verlag, New York, 1997.